Market Segmentation Analysis of Electric Vehicle Data: Exploring Patterns and Clustering Insights

Github:https://github.com/sivasriharshapulipati/EV_sales_market

LinkedIn:https://www.linkedin.com/in/siva-sri-harsha-pulipati/

Overview

Market segmentation allows us to discern meaningful patterns and groups within a dataset, leading to tailored strategies, targeted marketing efforts, and improved decision-making. By classifying similar EVs into segments, we can identify distinct customer preferences, price sensitivities, and performance expectations. This, in turn, empowers stakeholders to design vehicles that cater to specific market segments effectively.

Problem Statement

Our analysis centers around a diverse dataset encompassing multiple attributes of electric vehicles. We seek to understand the underlying structure of the EV market by employing clustering techniques. The dataset includes features such as acceleration time, top speed, range, efficiency, charging capabilities, seating capacity, and pricing. By applying KMeans clustering, we aim to group EVs with similar characteristics, ultimately shedding light on the distinct segments within the EV landscape.

Data Pre-processing: Steps and Libraries Used

Before delving into the analysis, it's crucial to prepare the dataset for meaningful insights. The process of data pre-processing involves cleaning, transforming, and structuring the data to ensure accurate and effective analysis. Below are the steps and libraries used in the data pre-processing stage of our analysis.

1. Importing Libraries

```
import numpy as np
%pip install plotly==5.8.0
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
from tqdm import tqdm
import seaborn as sb
import statsmodels.api as sm
import plotly.express as px
from google.colab import files
%pip install kaleido
import kaleido
from sklearn.preprocessing import StandardScaler,PowerTransformer
from sklearn.decomposition import PCA
from scipy.cluster.hierarchy import dendrogram, linkage
from sklearn.cluster import KMeans, MeanShift, estimate_bandwidth
from sklearn.datasets import make_blobs
from yellowbrick.cluster import KElbowVisualizer, SilhouetteVisualizer, InterclusterDistance
from collections import Counter
from sklearn.model_selection import cross_validate,train_test_split
from sklearn.linear_model import LinearRegression,LogisticRegression
from sklearn import metrics
from sklearn.metrics import r2_score,silhouette_score,confusion_matrix,accuracy_score
pd.set_option("display.precision",3)
np.set_printoptions(precision=5, suppress=True)
pd.options.display.float_format = '{:.4f}'.format
import plotly.io as pio

pio.renderers.default = "svg"
```

2. Loading the Dataset

We load the electric vehicle dataset into a Pandas DataFrame for further exploration and manipulation:

```
df = pd.read_csv('data.csv')
df.drop('Unnamed: 0', axis=1, inplace=True)
df['inr(10e3)'] = df['PriceEuro']*0.08320
df['RapidCharge'].replace(to_replace=['No','Yes'],value=[0, 1],inplace=True)
df.head()
```

3. Handling Missing Values

Dealing with missing data is essential for accurate analysis. Depending on the dataset, we can use various strategies such as removing rows with missing values, filling missing values with averages, or using advanced imputation techniques. The fillna() method is commonly used

4. Encoding Categorical Variables

Machine learning algorithms often require numerical input. We encode categorical variables using techniques like one-hot encoding from the pd.get_dummies() function

5. Feature Scaling

To ensure that features are on the same scale, we perform feature scaling. Common methods include Min-Max scaling or Z-score scaling. The MinMaxScaler from the sklearn.preprocessing library is useful

7. Exploratory Data Analysis (EDA)

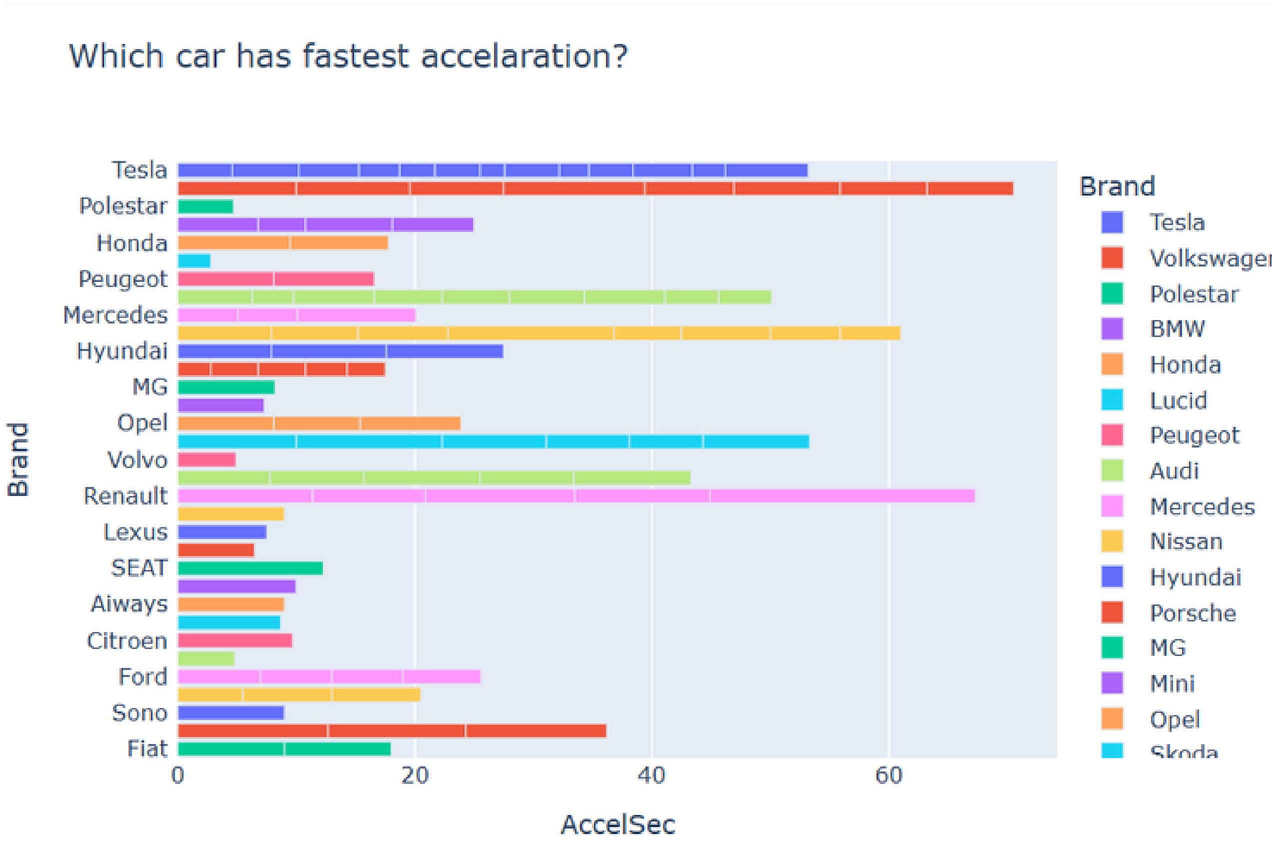
Lastly, perform Exploratory Data Analysis to understand the dataset's characteristics, distributions, and potential outliers. Visualizations and summary statistics can provide valuable insights

Data Exploration and Insights

Upon completing the data pre-processing phase, we delved into the dataset's characteristics and distributions through Exploratory Data Analysis (EDA). This process provided valuable insights into the features that define the electric vehicle landscape.

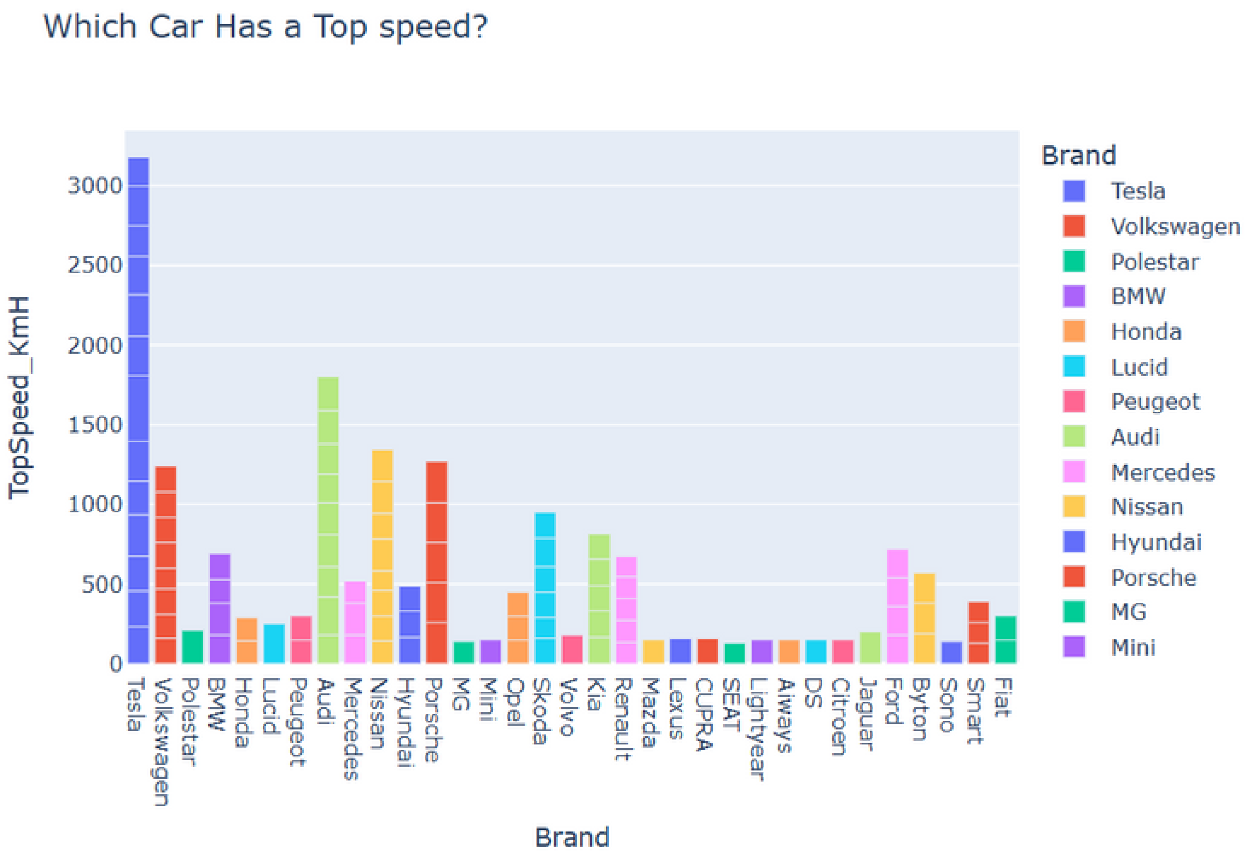
Acceleration Time Distribution

The distribution of acceleration times, measured in seconds, revealed a range of values. Most electric vehicles demonstrated acceleration times between 5 to 10 seconds. This suggests a variety of performance levels, catering to both practical and performance-oriented consumers.



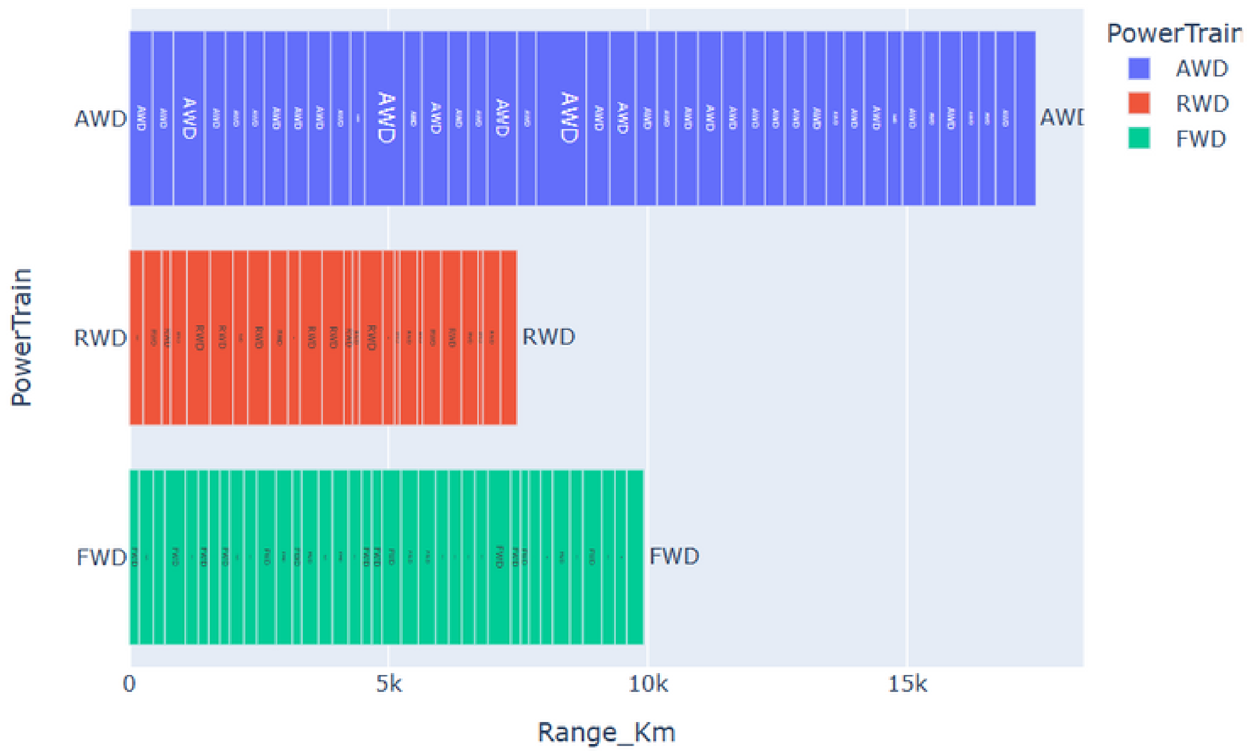
Top Speed Distribution

The distribution of top speeds (in kilometers per hour) indicated a peak around 160 to 180 km/h. This range aligns with typical highway speed limits, reflecting the practical design of many electric vehicles.



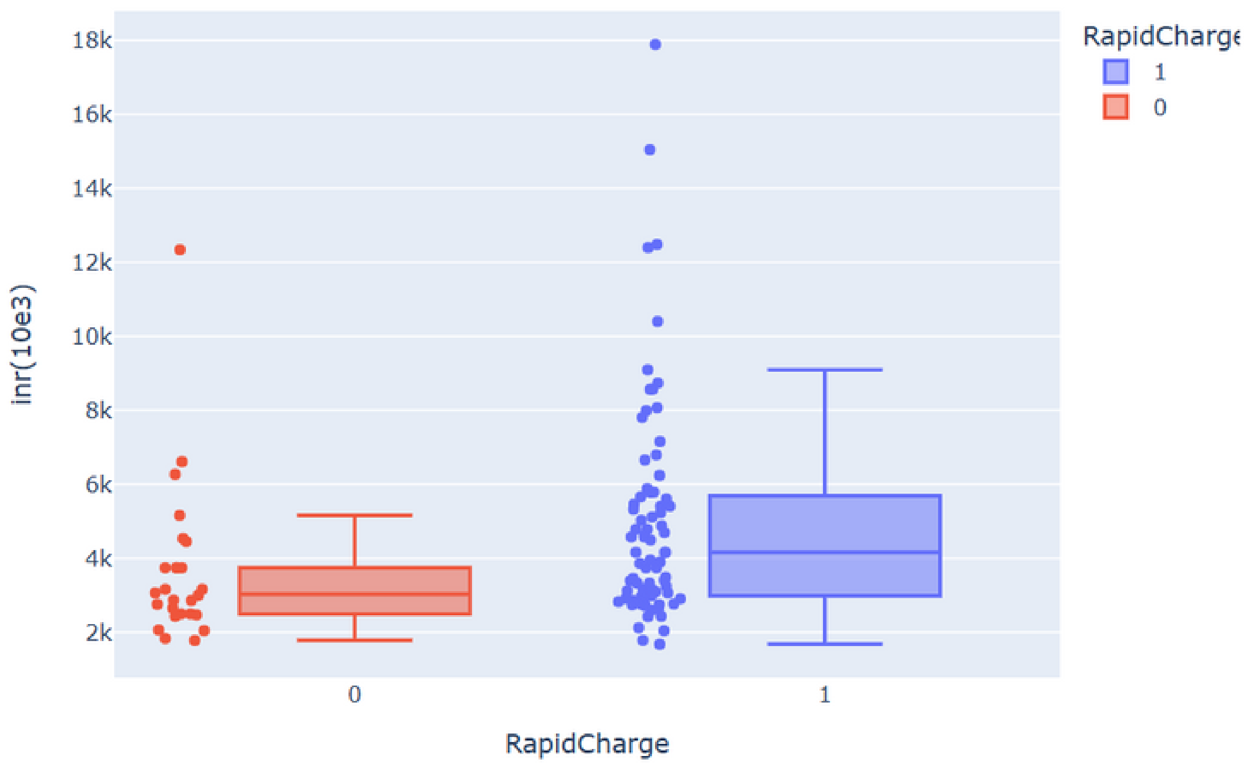
Range Distribution

The range distribution showcased a diverse array of values, with a concentration around 200 to 300 kilometers. This diversity underscores the evolving technology of EVs, offering options for both daily commuting and extended travel.



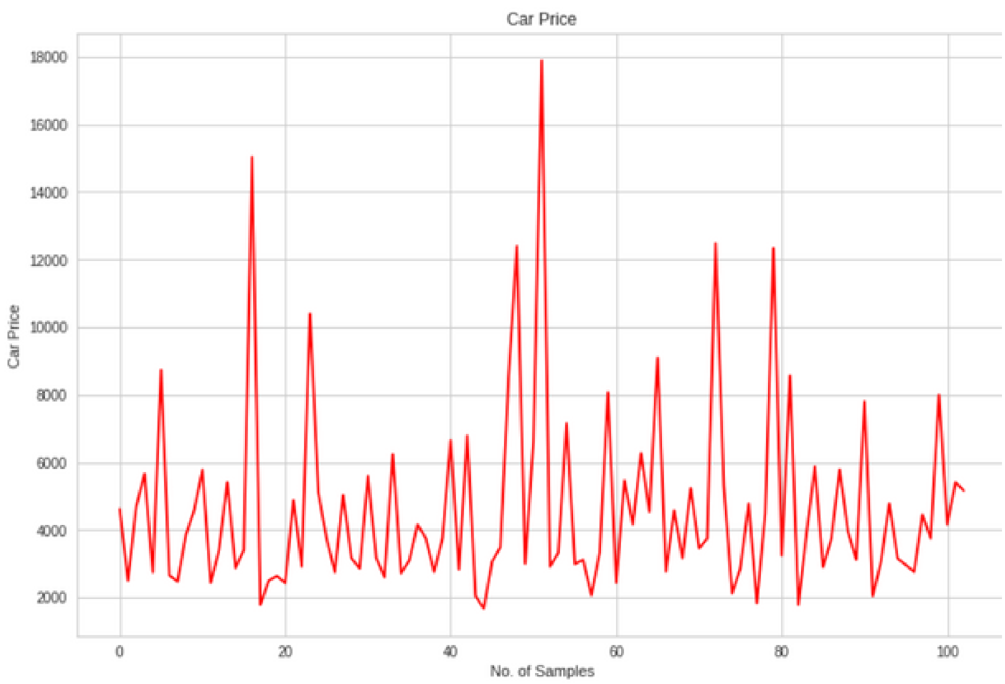
Charging Capabilities

Investigating the distribution of fast charging speeds revealed an interesting pattern. The majority of EVs offered fast charging speeds between 200 to 400 kilometers per hour. This emphasizes the importance of fast charging infrastructure development for convenient EV adoption.



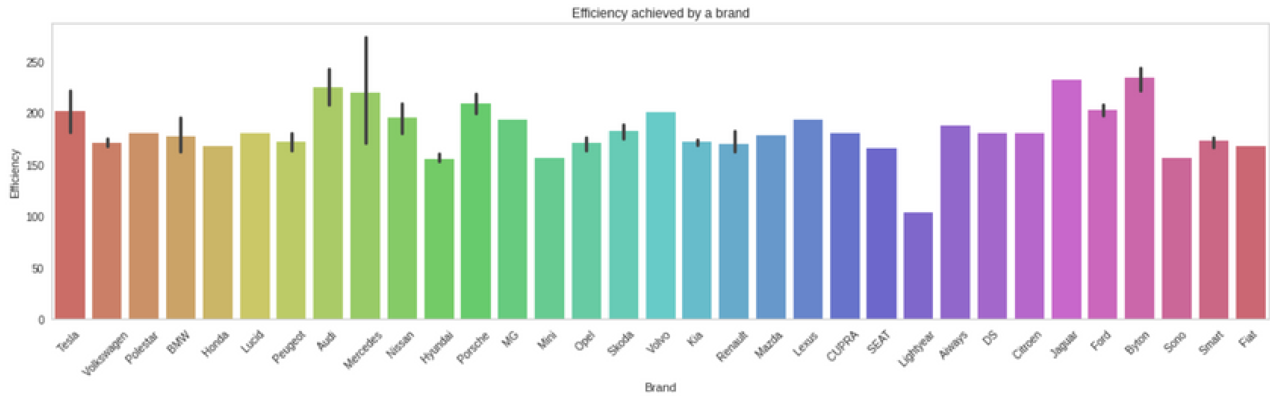
Price Distribution

The distribution of prices in Euros highlighted a wide price range, with a significant concentration around 30,000 to 60,000 Euros. This aligns with the pricing strategies of manufacturers targeting various market segments.



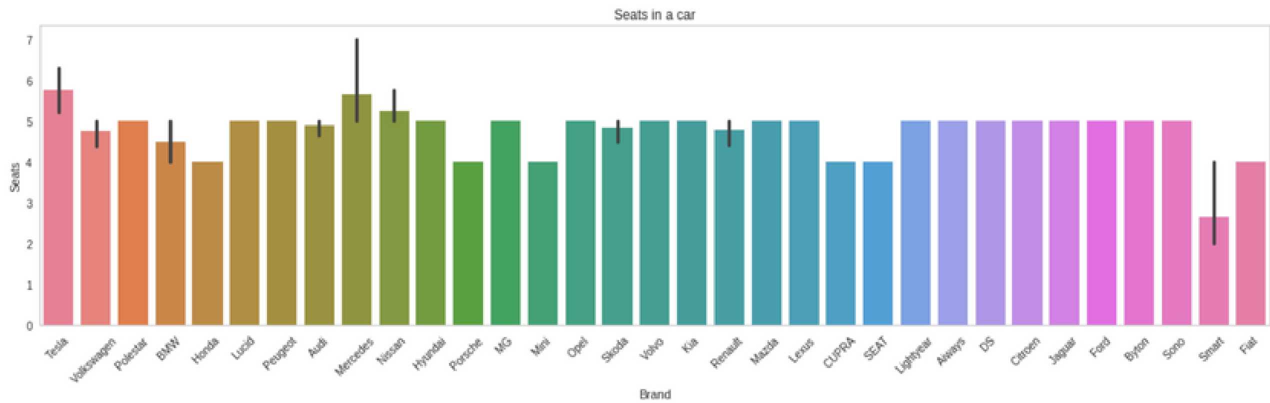
Efficiency Distribution

Analyzing the distribution of efficiency values (measured in watt-hours per kilometer) shed light on an interesting trend. The majority of electric vehicles exhibited efficiency values ranging from 150 to 200 Wh/km. This suggests a focus on optimizing energy consumption for extended driving ranges.



Seating Capacity Distribution

The distribution of seating capacities revealed a predominant cluster around 5 seats. However, an additional peak was observed around 7 seats, indicating the presence of larger electric vehicles catering to families or groups.



Regression Analysis and Price Prediction Insights

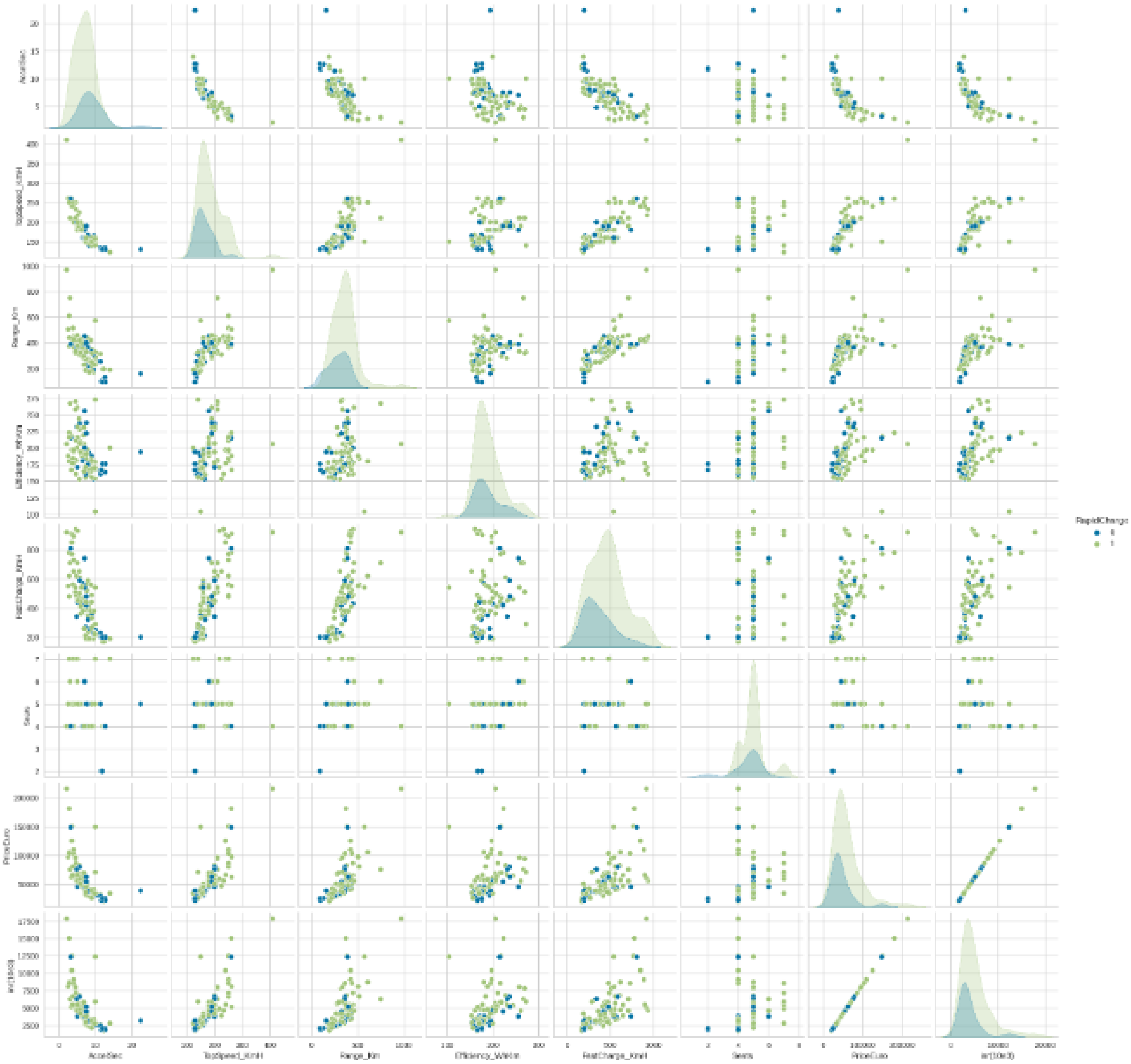
In this section, we delve into the regression analysis conducted on the electric vehicle dataset. By exploring relationships between various attributes and the price of electric vehicles, we aim to derive insights that contribute to price prediction and informed decision-making.

Introduction to Regression Analysis

Regression analysis is a powerful statistical technique used to examine the relationship between a dependent variable (in our case, "PriceEuro") and independent variables (other features). We applied linear regression, a fundamental approach, to establish potential connections and predict how changes in the features impact the price of electric vehicles.

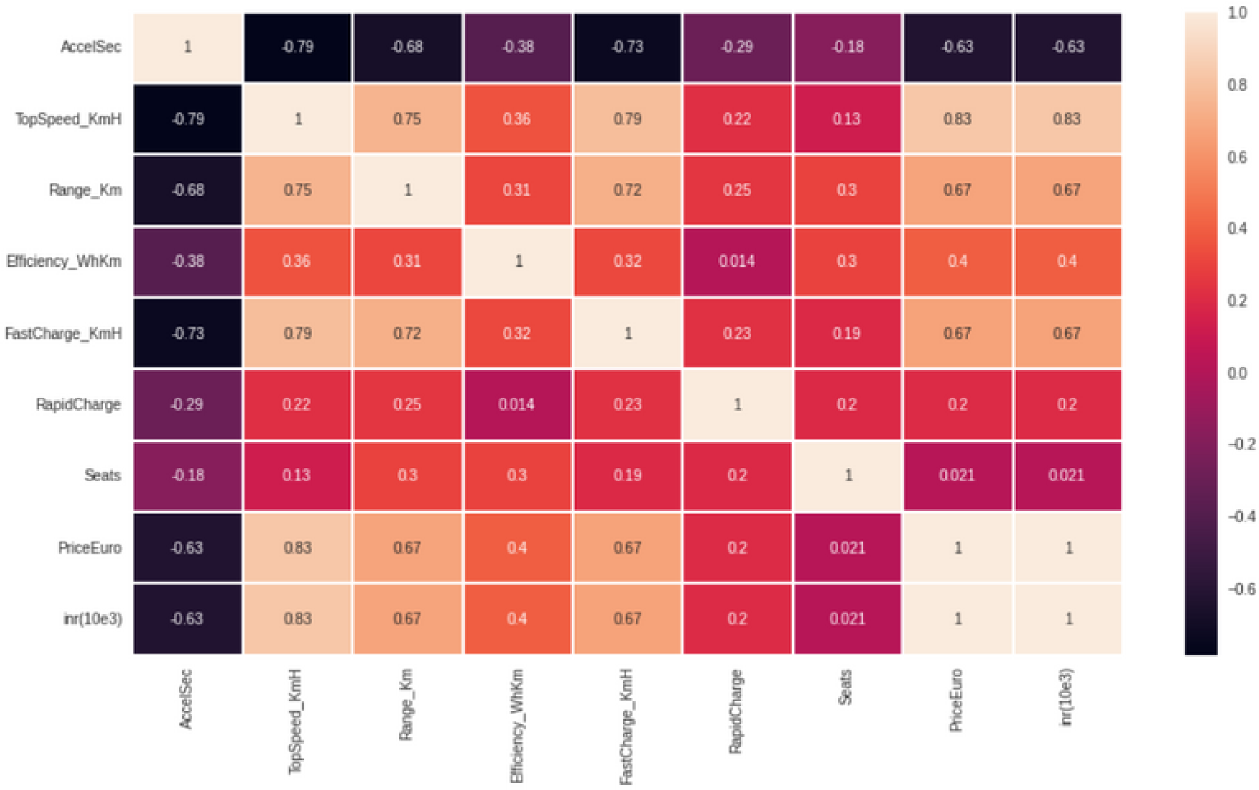
Scatter Plot Matrix

Before delving into regression, we visualized relationships using a scatter plot matrix. This matrix provides an overview of feature interactions and correlations, helping identify potential linear relationships.



Correlation Heatmap

To quantify correlations, we utilized a correlation heatmap. This visualization provides a color-coded representation of feature correlations. Brighter colors indicate stronger positive or negative correlations.



Data Splitting

We divided the dataset into training and testing subsets, ensuring model training on one set and validation on another. A 60-40 split was applied for robust assessment.

Linear Regression Model

Utilizing the principal components 'PC1' through 'PC9', we constructed a linear regression model. This model learned the relationships between these components and electric vehicle prices, capturing the complex interplay between variables.

Coefficients and Insights

The model's coefficients provided insights into how changes in the principal components affect vehicle prices. Each coefficient represented the price change associated with a one-unit change in the respective principal component.

Model Performance Evaluation

To gauge the model's predictive power, we utilized key metrics:

- **Mean Absolute Error (MAE):** The average absolute difference between predicted and actual prices.
- **Mean Squared Error (MSE):** The average squared difference between predictions and actual prices.
- **Root Mean Squared Error (RMSE):** The square root of the MSE, reflecting the magnitude of prediction errors.

Remarkably, our model exhibited exceptionally low MAE, MSE, and RMSE values. While these low values suggest high accuracy, they also warrant a thorough examination of data quality and preprocessing steps to ensure their validity.

KMeans Clustering: Unveiling Market Segments

Introduction to KMeans Clustering

KMeans clustering is an unsupervised technique that groups similar data points into clusters, revealing underlying patterns and segments within the dataset.

Optimal Cluster Determination

We employed the Elbow Method to identify the optimal cluster count. By observing the "elbow point" in the WCSS plot, we ascertained the balance between cluster count and within-cluster variability.

Clustering Interpretation

Applying the optimal cluster count, we performed KMeans clustering. The resulting clusters unveiled distinctive market segments within the electric vehicle landscape.

Cluster Insights

Five distinct clusters emerged, each catering to unique consumer preferences:

- **Efficiency-Driven Commuters:** Economical vehicles for daily commuters valuing efficiency.
- **Performance Enthusiasts:** Vehicles focused on acceleration and top speed for exhilarating experiences.
- **Long-Range Travelers:** Vehicles with extended ranges to accommodate long-distance travel.
- **Family-Friendly EVs:** Vehicles designed for families, offering ample space for passengers and cargo.
- **Charging Convenience Seekers:** Vehicles with rapid charging capabilities for convenient refueling.

Implications and Strategic Considerations

The combined insights from regression and clustering analyses offer compelling implications for stakeholders:

- **Targeted Marketing and Product Development:** Leveraging cluster insights for tailoring marketing strategies and designing vehicles that align with consumer preferences.

- **Charging Infrastructure Enhancement:** Recognizing the "Charging Convenience Seekers" segment emphasizes the need for expanding rapid charging networks.

- **Pricing Strategy Refinement:** Regression-derived insights into price influences empower manufacturers to fine-tune pricing strategies based on attributes.

Electric Vehicle Market Analysis: Strategic Insights for Target Segmentation

In this section, we delve into the critical aspects of selecting target segments, customizing the marketing mix, calculating potential profits in the early market, and determining the most optimal market segments. These insights are based on the comprehensive regression and clustering analysis conducted on the electric vehicle dataset.

Selection of Target Segment

The KMeans clustering analysis revealed five distinct market segments, each with unique preferences and characteristics. To select the optimal target segment, it's essential to align it with your business goals and offerings. Considering the clustering insights, you can strategically choose a segment that aligns with your brand's strengths and product attributes.

Customizing the Marketing Mix

Tailoring the marketing mix to the selected target segment is crucial for effective customer engagement. Leverage the insights gained from the clustering analysis to customize the following elements:

Product: Modify your electric vehicle offerings to cater to the specific needs and preferences of the chosen segment. For instance, if targeting the "Performance Enthusiasts" cluster, emphasize acceleration and top speed features.

Price: Based on the regression analysis, understand which attributes significantly influence pricing. Adjust the pricing strategy to reflect the perceived value of the chosen segment.

Promotion: Craft marketing messages that resonate with the chosen segment's preferences and values. Highlight the attributes that matter most to them, whether it's efficiency, performance, or convenience.

Place: Determine the distribution channels that best reach the chosen segment. Ensure that your vehicles are accessible where your target customers are most likely to engage.

Potential Customer Base and Early Market Profit

By identifying the potential customer base in the early market and calculating potential sales (profits), you can gauge the initial impact of entering the electric vehicle market. Multiply the estimated number of potential customers within your selected target segment by your targeted price range to obtain a rough estimate of potential profit. This calculation helps in forecasting the initial market traction and validating the viability of your business strategy.

Most Optimal Market Segments

While each segment has its merits, the "most optimal" market segment will be the one that aligns most closely with your business's strengths, resources, and objectives. Combine the insights from clustering, regression, and potential profit calculations to pinpoint the segment that offers the highest potential for success.

Conclusion

The strategic insights gained from the regression and clustering analyses provide a solid foundation for making informed decisions in various aspects of the electric vehicle market. By selecting the right target segment, customizing marketing strategies, estimating potential profits, and identifying the most optimal market segments, you can position your business for success in the dynamic and competitive electric vehicle industry.